CRC（続き）

情報数理学特別講義I 第6回

2025-11-05

今回の講義内容

- CRCのパリティ検査行列
- CRCの設計
- 2元対称通信路(BSC)
- 符号の重み分布と重み分布母関数
- 重み分布母関数を用いた誤り見逃し確率の計算
- MacWilliamsの恒等式

CRCは線形符号

CRCのパリティ検査行列は次のように定義できる.

$$H = \left(Rem(x^0)_{g(x)}, Rem(x^1)_{g(x)}, \ldots, Rem(x^{n-1})_{g(x)}\right)$$

$Rem(x^i)_{g(x)}$ は多項式だが，上記では２元の列ベクトルとして扱っている.

例えば先の生成多項式 $g(x) = 1 + x + x^4$ で符号化したCRCのパリティ検査行列は

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

となり，受信語（ベクトル）との積をとることで，誤り検出が可能となる.

計算例

先の H の定義とは高次と低次が逆順になっていることに注意.

```
c = [1,1,0,1,0,1,0,1,0,0,1,1] # x^4(x^7+x^6+x^4+x^2+1) + Rem(x^4 m(x)/g(x))

H = [0 1 0 1 1 0 0 1 0 0 0 1;
     1 1 1 0 1 0 1 1 0 0 1 0;
     1 1 0 1 0 1 1 0 0 1 0 0;
     1 0 1 0 1 1 0 0 1 0 0 0]
julia> H*c .% 2
4-element Vector{Int64}:
 0
 0
 0
 0
```

計算例

```
r = [1,1,0,1,0,1,0,1,0,0,1,0] # x^4(x^7+x^6+x^4+x^2+1) + x
julia> H*r .% 2
4-element Vector{Int64}:
 1
 0
 0
 0
 # LFSRで計算した [1, 0, 0, 0] と同じ結果
```

$r(x) = x^4(x^7 + x^6 + x^4 + x^2 + 1) + x$
$c(x) = x^4(x^7 + x^6 + x^4 + x^2 + 1) + x + 1$
であることから,
$e(x) = 1$. 得られたシンドロームも $Rem(e(x)/g(x))$ となっている.
また以上より, **多項式の剰余という操作は線形写像である**ことが実感できるだろう.

CRCの設計

CRCの誤り検出能力は生成多項式によって決定される.
以降は生成多項式をどのように設計するかについて考える.

CRCが考慮すべき誤りモデル

CRCによる誤り検出を考える際，一般に２つの誤りを考える．

- バースト誤り：誤り位置が集中する（ある箇所がほぼ連続して誤る）誤り
  - $\underline{e} = 000000000000000000000000000010110111000$
- ランダム誤り：誤り位置がランダムな誤り
  - $\underline{e} = 000100000000000000000000000010000000$

r 次の生成多項式を持つCRCは，r ビット以下のバースト誤りを1つ検出できる．
→ 保護すべきバースト誤りの長さにより生成多項式の次数 r が決まる．
ランダム誤りだけを想定する場合は，システムが許容できるパリティビット数で生成多項式の次数が決まることが多い（生成多項式の次数 = パリティビット数）．

例えば

あるシステムにおいて，用いるデータの信頼性を高めるため256bitのデータに対して32bitのCRC(CRC-32)を付加する必要がある．このCRC-32の設計を依頼されたらどうするか？

よくあるパタン：
ググる → Wikipediaにたどり着く → CRC-32の生成多項式が沢山ある →
標準的（と思われる）ものを選択する

※右表は Wikipedia:Cyclic redundancy check から抜粋

CRC-16-Chakravarty	Optimal for payloads ≤64 bits[29]	0x2F15	0xA8F4	0x51E9	0x978A
CRC-16-ARINC	ACARS applications[45]	0xA02B	0xD405	0xA80B	0xD015
CRC-16-CCITT	X.25, V.41, HDLC *FCS*, XMODEM, Bluetooth, PACTOR, SD, DigRF, many others; known as *CRC-CCITT*	0x1021	0x8408	0x811	0x8810[11]
		$x^{16} + x^{12} + x^5 + 1$			
CRC-16-CDMA2000	mobile networks[26]	0xC867	0xE613	0xCC27	0xE433
CRC-16-DECT	cordless telephones[46]	0x0589	0x91A0	0x2341	0x82C4
		$x^{16} + x^{10} + x^8 + x^7 + x^3 + 1$			
CRC-16-T10-DIF	SCSI DIF	0x8BB7[47]	0xEDD1	0xDBA3	0xC5DB
		$x^{16} + x^{15} + x^{11} + x^9 + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1$			
CRC-16-DNP	DNP, IEC 870, M-Bus	0x3D65	0xA6BC	0x4D79	0x9EB2
		$x^{16} + x^{13} + x^{12} + x^{11} + x^{10} + x^8 + x^6 + x^5 + x^2 + 1$			
CRC-16-IBM	Bisync, Modbus, USB, ANSI X3.28, SIA DC-07, many others; also known as *CRC-16* and *CRC-16-ANSI*	0x8005	0xA001	0x4003	0xC002
		$x^{16} + x^{15} + x^2 + 1$			
CRC-16-OpenSafety-A	safety fieldbus[33]	0x5935	0xAC9A	0x5935	0xAC9A[11]
CRC-16-OpenSafety-B	safety fieldbus[33]	0x755B	0xDAAE	0xB55D	0xBAAD[11]
CRC-16-Profibus	fieldbus networks[48]	0x1DCF	0xF3B8	0xE771	0x8EE7

CRCの目的

正しくCRCの生成多項式を選択するには，原理原則を理解する必要がある．
CRCの目的はなにか？

CRCの誤り検出能力

CRCの目的は受信語に含まれる誤りを検出すること.
CRCの誤り検出能力は受信語に誤りが含まれているにもかかわらず,
誤りを検出できない確率（undetected error rate: UDER）で評価される.
UDERは以後，誤り見逃し率と書く.

CRCが誤りを検出できない事象

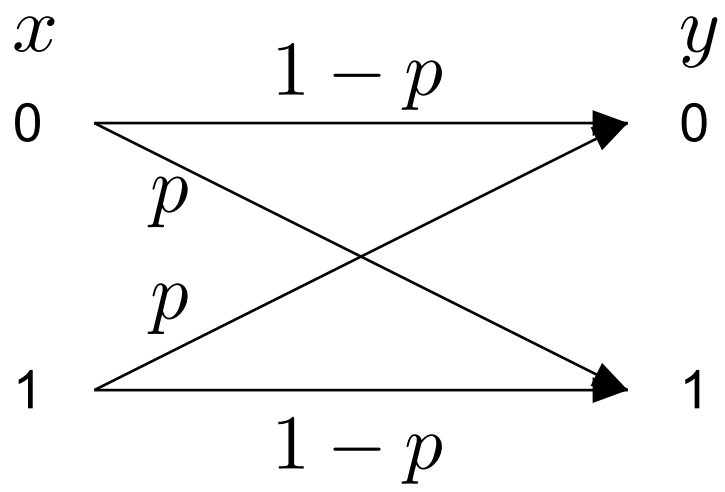
CRCは生成多項式による剰余の結果が非ゼロのときに誤りを検出する.
つまり，誤り多項式 $e(x)$ が生成多項式 $g(x)$ で割り切れるときに誤りを検出できない.

- $Rem(e(x))_{g(x)} = 0$ となる $e(x)$ が発生する確率がUDER.

以下では2元対称通信路(BSC: Binary Symmetric Channel)でのUDERを考える.

2元対称通信路(BSC)

2元対称通信路(BSC)は，送信ビットが反転する確率が p である通信路である．



BSCでは各ビットの反転は独立に発生すると仮定する．
nビット送信し，kビット誤りが発生する確率は二項分布に従う．

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

符号の重み分布

重みとは，符号語に含まれる１の数.
符号の重み分布とは，ある符号が持つ重み i の符号語の数 A_i の分布.

例えば長さ3の単一パリティ検査符号 $\mathcal{C} = \{(000), (101), (011), (110)\}$ の重み分布は

$$A_0 = 1, A_2 = 3.$$

重み分布を係数に持つ多項式を重み分布母関数と呼び,

$$A(z) = \sum_{i=0}^{n} A_i z^i$$

と定義する．上の例では $A(z) = 1 + 3z^2$ となる.

符号の最小ハミング距離と最小重み（再掲）

$GF(2)$ 上のベクトル $\underline{u}, \underline{v}$ 間のハミング距離 $d_H(\underline{u}, \underline{v})$ は ベクトル $\underline{w} = \underline{u} + \underline{v}$ の重み $wt(\underline{w})$ と一致する.

- ベクトル $\underline{w}$ の i 番目の要素 w_i が 1 のとき，つまり $w_i = 1$ は $u_i \neq v_i$ の場合だから.

線形符号では符号語の和も符号語である.

- $\underline{c}_3 = \underline{c}_1 + \underline{c}_2$ のとき，$H\underline{c}_3 = H(\underline{c}_1 + \underline{c}_2) = H\underline{c}_1 + H\underline{c}_2 = \underline{0} + \underline{0} = \underline{0}$ となる.

つまり，符号語間の相違を表すベクトル自体が符号語となるため，**符号語の重みを調べることと符号語間のハミング距離を調べることは同値**である.

よって，線形符号の最小ハミング距離は，符号語の最小重みを調べることで求まる.

CRCのBSC上での誤り見逃し率

反転確率 p のBSCにおける誤り見逃し率 P_{ud} は以下で計算される.

$$P_{ud} = \sum_{i=1}^{n} A_i p^i (1-p)^{n-i}.$$

ただし，A_i は重み i の符号語数（重み分布）．また,

$$\begin{aligned} P_{ud} &= \sum_{i=1}^{n} A_i p^i (1-p)^{n-i} = (1-p)^n \sum_{i=1}^{n} A_i \left(\frac{p}{1-p} \right)^i \\ &= (1-p)^n \left(A\left(\frac{p}{1-p} \right) - 1 \right) \end{aligned}$$

のように式変形することで，重み分布母関数を使って書くことができる.

→ **重み分布は符号の性能を決定する重要な指標.**

誤り見逃し率の上限

反転確率 $p = \frac{1}{2}$ のとき

$$P_{ud} = \frac{1}{2^n} \sum_{i=1}^{n} A_i = \frac{2^k - 1}{2^n} \simeq \frac{1}{2^r}.$$

※ $\sum_{i=1}^{n} A_i$ は全零符号語以外の符号語の和なので $2^k - 1$ となる．r はCRCのパリティビット数 $r = n - k$．反転確率 p が大きい状況では，一様に発生した剰余多項式がたまたま 0 になる確率と考えても良い．

例：CRC-32では $1/2^{32} = 2.3 \times 10^{-10}$, つまり100億分の2.

符号の重み分布の導出

符号の重み分布は，符号の生成多項式によって決定される．
解析式が与えられている場合は重み分布を計算することができるが，
一般には重み分布の解析式が与えられていない．

解析式が与えられていない場合，符号語の重み分布を求めるには，

- 符号の生成多項式を用いて重み分布を計算する方法
  - 2^k個の相異なるメッセージを符号化して符号語を生成する．
    $k > 32$ など，メッセージ数が多い場合は計算が困難．
- 符号の**双対符号**の重み分布を用いて重み分布を計算する方法．

双対符号（dual code）

符号 $\mathcal{C}$ の双対符号 $\mathcal{C}^{\perp}$ は以下のように定義される.

$$\mathcal{C}^{\perp} = \{\underline{v} \in GF(2)^n | \underline{v} \cdot \underline{c} = 0, \forall \underline{c} \in \mathcal{C}\}.$$

つまり，$\mathcal{C}^{\perp}$ は $\mathcal{C}$ の符号語と直交するベクトルの集合.

符号 $\mathcal{C}$ の次元が k の時，$\mathcal{C}^{\perp}$ の次元は $n-k$ となる.

双対な符号同士の性質

符号 $\mathcal{C}$ の生成行列が G，パリティ検査行列が H の時,
双対符号 $\mathcal{C}^{\perp}$ の生成行列 $G^{\perp}$ とパリティ検査行列 $H^{\perp}$ はそれぞれ

$$G^{\perp} = H,$$
$$H^{\perp} = G.$$

$GH^T = \mathbf{0}$ となること,
G の各行ベクトルは符号 $\mathcal{C}$ の符号語であること,
H の各行ベクトルは符号 $\mathcal{C}$ の符号語と直交するベクトル（内積が0）であること
から明らか.

符号 $\mathcal{C}$ の双対符号 $\mathcal{C}^{\perp}$ 間の重み分布の関係

符号 $\mathcal{C}$ の重み分布母関数を $A(z)$，双対符号 $\mathcal{C}^{\perp}$ の重み分布母関数を $B(z)$ とする時，$A(z)$ と $B(z)$ の関係は以下のようになる（MacWilliamsの恒等式）.

$$A(z) = 2^{-r}(1+z)^{n} B\left(\frac{1-z}{1+z}\right).$$

ただし，r は双対符号 $\mathcal{C}^{\perp}$ の次元.

MacWilliamsの恒等式を用いると，双対符号の重み分布母関数 $B(z)$ から，符号の重み分布母関数 $A(z)$ 求めることができる．n や k が大きな符号であっても，$r = n - k$ が小さいことは多いので，その際には $B(z)$ を求めた後に MacWilliamsの恒等式により $A(z)$ を求める.

※$A(z)$ と $B(z)$ は対称関係のため，$A(z)$ から $B(z)$ を求めることもできる.

MacWilliamsの恒等式を用いた重み分布母関数の計算

単一パリティ検査符号の双対符号は繰り返し符号である.
繰り返し符号の重み分布母関数は $B(z) = 1 + z^n$ である.
単一パリティ検査符号の重み分布母関数は
n が偶数の時

$$A(z) = 1 + \binom{n}{2} z^2 + \binom{n}{4} z^4 + \cdots + \binom{n}{n} z^n$$

n が奇数の時

$$A(z) = 1 + \binom{n}{2} z^2 + \binom{n}{4} z^4 + \cdots + \binom{n}{n-1} z^{n-1}$$

となる. 単一パリティ検査符号は偶重み語をすべて持つ.

MacWilliamsの恒等式を用いた重み分布母関数の計算(続)

MacWilliams恒等式によれば

$$
\begin{aligned}
A(z) &= 2^{-1}(1+z)^n B\left(\frac{1-z}{1+z}\right) \\
&= 2^{-1}(1+z)^n \left(1+\left(\frac{1-z}{1+z}\right)^n\right) \\
&= 2^{-1}\left((1+z)^n + (1-z)^n\right) \\
&= 2^{-1}\sum_{w=0}^{n}\left(\binom{n}{w} + (-1)^w \binom{n}{w}\right) z^w
\end{aligned}
$$

が得られる．最後の式は偶重み語のみを持つ母関数となっていることから，先に述べた通り単一パリティ検査符号の重み母関数と一致する．

MacWilliamsの恒等式を用いた誤り見逃し確率の計算

$$
\begin{aligned}
P_{ud} &= (1-p)^n \left(A\left(\frac{p}{1-p}\right) - 1 \right) \\
&= (1-p)^n \left(2^{-(n-k)}(1+\frac{p}{1-p})^n B\left(\frac{1-\frac{p}{1-p}}{1+\frac{p}{1-p}}\right) - 1 \right) \\
&= (1-p)^n \left(2^{-(n-k)}(\frac{1}{1-p})^n B\left(1-2p\right) - 1 \right) \\
&= 2^{-(n-k)} B\left(1-2p\right) - (1-p)^n
\end{aligned}
$$

となり，$B(z)$ が求まれば誤り見逃し確率 P_{ud} を計算できる.

巡回符号の例

符号長21の2元巡回符号を考えよう．$x^{21}-1$ を$GF(2)$上で因数分解すると，

$$\begin{aligned} x^{21}-1 =& (x-1)\left(x^6+x^4+x^2+x+1\right) \\ & \times \left(x^3+x^2+1\right)\left(x^6+x^5+x^4+x^2+1\right) \\ & \times \left(x^2+x+1\right)\left(x^3+x+1\right) \end{aligned}$$

となるので，生成多項式

$$g_1(x) = \left(x^6+x^4+x^2+x+1\right)\left(x^3+x^2+1\right)$$

によって，2元 [21,12] 巡回符号を得る．また，生成多項式

$$g_2(x) = \left(x^6+x^4+x^2+x+1\right)\left(x^3+x+1\right)$$

によっても，2元 [21,12] 巡回符号を得る．このように同一の $[n,k]$ を有する巡回符号は複数存在するので，実際にはこれらの中から最小距離の大きなものを利用する．

生成多項式 $g_1(x), g_2(x)$ の重み分布

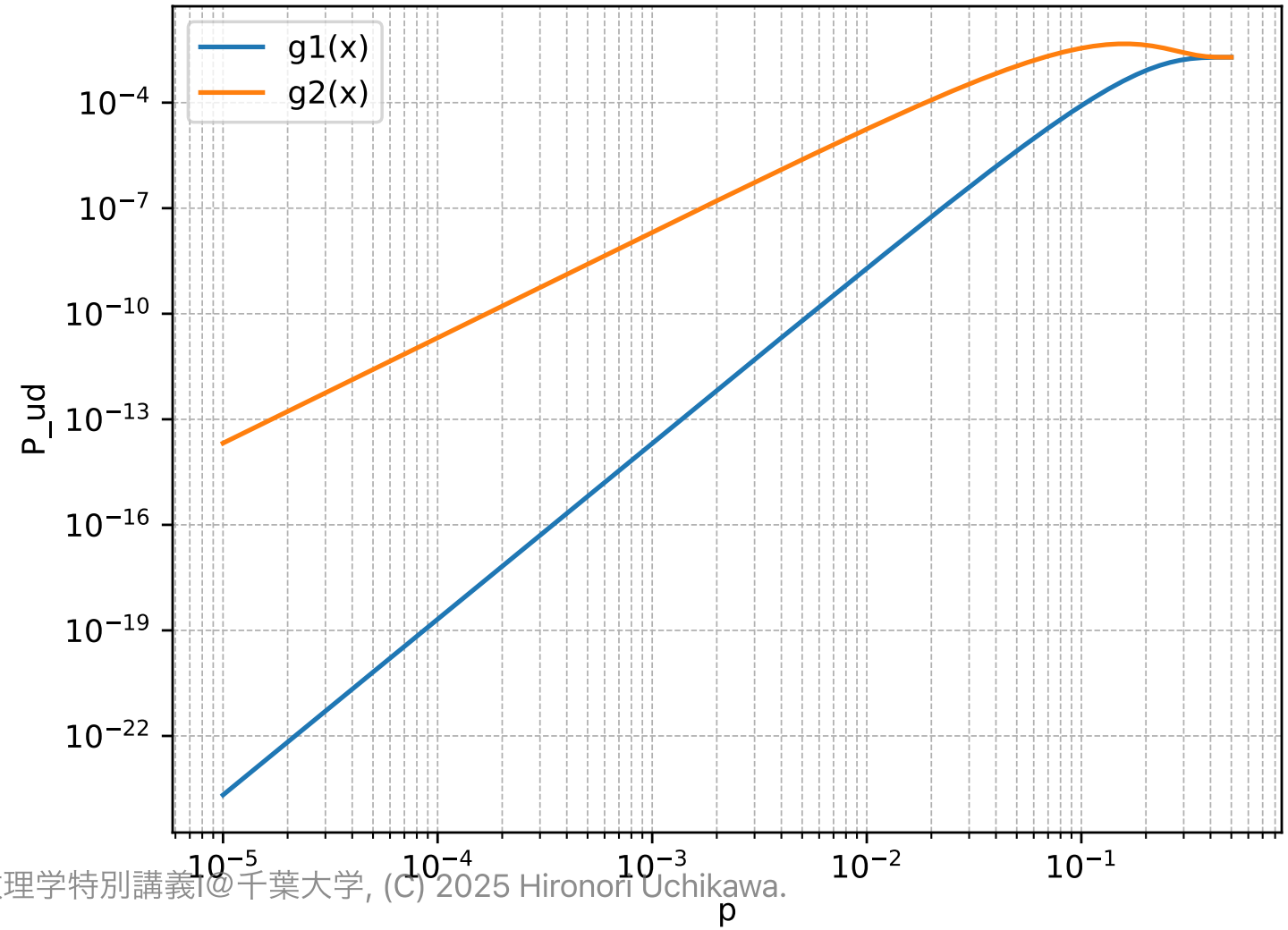
$g_1(x)$

$A_0 = 1, A_5 = 21, A_6 = 168, A_7 = 360, A_8 = 210, A_9 = 280, A_{10} = 1008, \ldots$

$g_2(x)$

$A_0 = 1, A_3 = 21, A_4 = 21, A_6 = 147, A_7 = 297, A_8 = 147, A_9 = 343, \ldots$

生成多項式 $g_1(x), g_2(x)$ のCRCとしての性能



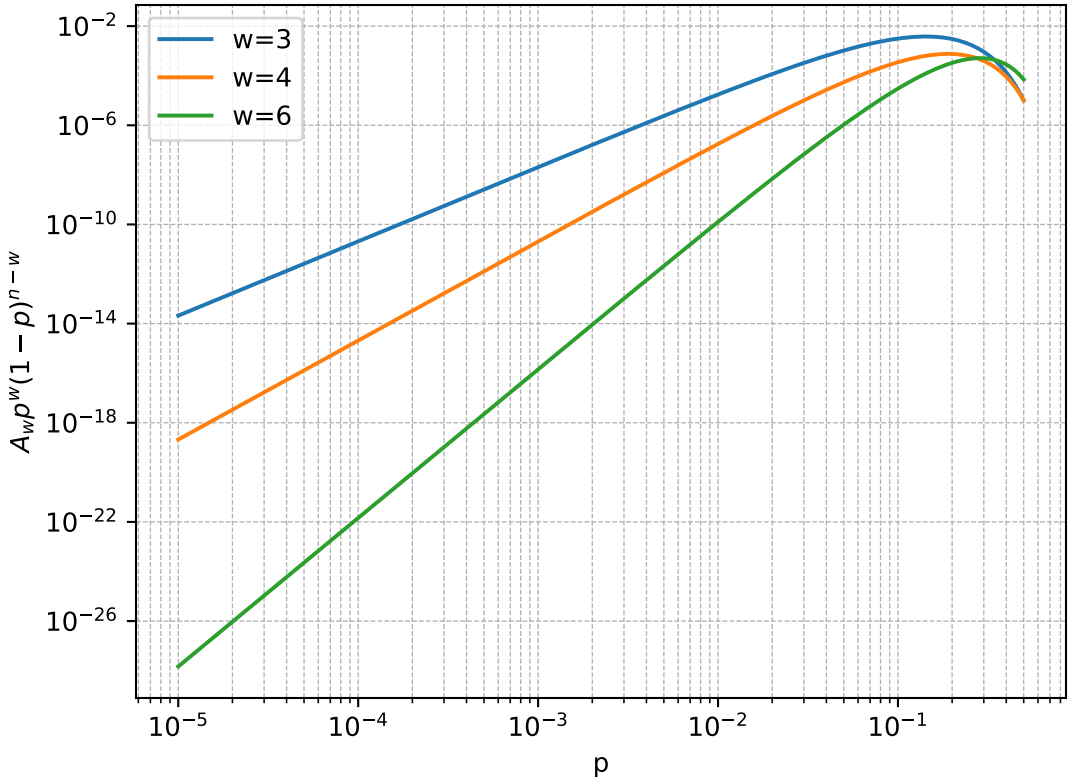
$g_2(x)$の誤り見逃し率について

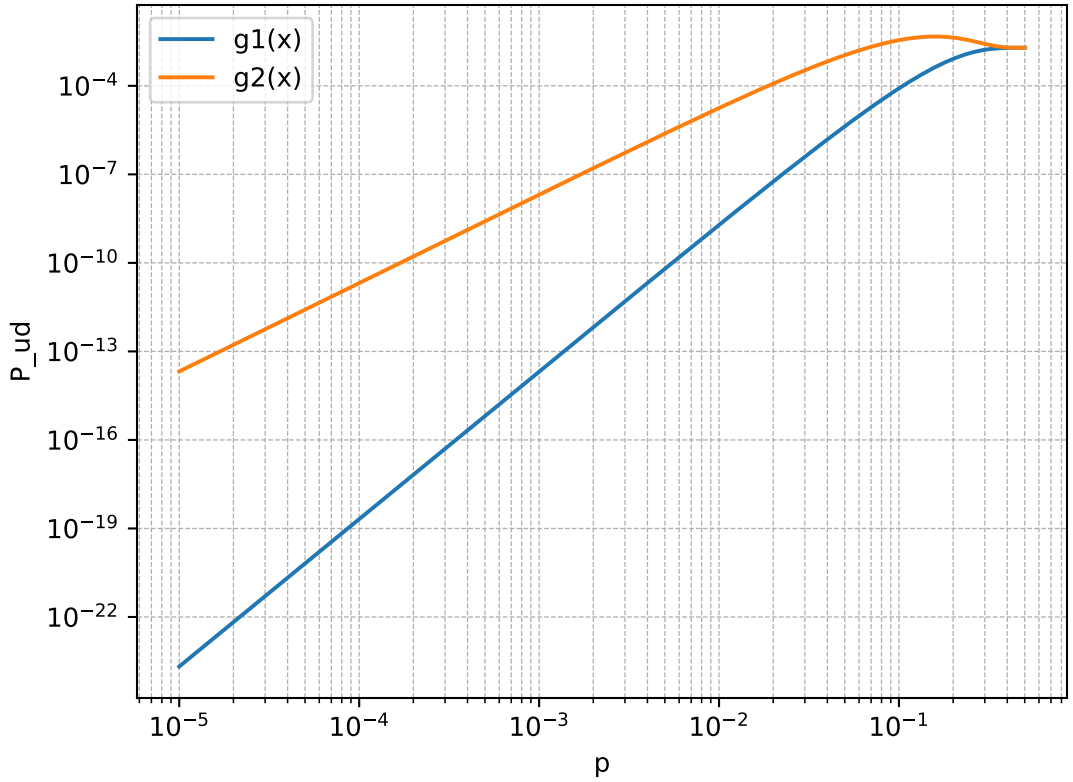
- $g_2(x)$ で構成した長さ$N = 21$のCRCは重み分布 A_w は低重み（$w = 3, 4$)を含んでいる.
- 誤り見逃し率の計算指揮に含まれる $A_w p^w (1-p)^{N-w}$ が $w = 3$のピークが$w/N \approx 0.14$, $w = 4$なら≈ 0.19でピークを持ち，$\frac{1}{2^r}$を超えるため，誤り見逃し率が単峰性を持つ.
- 各項は$A_w p^w (1-p)^{N-w}$の導関数

$$\frac{\mathrm{d}}{\mathrm{d}p}\left(A_w p^w (1-p)^{N-w}\right) = A_w p^{w-1} (1-p)^{N-w-1}(w - Np)$$

  が示す通り，$p = w/N$で極大を取り，$p < w/N$では増加，$p > w/N$では減少に転じる.

左は$g_2(x)$の$w = 3, 4, 6$の項の寄与，右は$g_1(x)$と$g_2(x)$の誤り見逃し率





今回の講義のまとめ

- CRCのパリティ検査行列
- CRCの設計
- 2元対称通信路(BSC)
- 符号の重み分布と重み分布母関数
- 重み分布母関数を用いた誤り見逃し確率の計算
- MacWilliamsの恒等式

演習

$GF(2)$上の符号長15の巡回符号について次の問いに答えよ.

1. $x^{15}-1$を$GF(2)$上の既約多項式の積に因数分解せよ.
2. 巡回符号は全部でいくつあるか.
3. 次元11の巡回符号の生成多項式を全て求めよ.
   次元11の巡回符号とはメッセージ長が11の [15,11]巡回符号のこと.
4. 巡回符号が構成可能な次元を全て求めよ.
   （ヒント）構成可能な生成多項式の次数を考えればよい.